DOCUMENT RESUME

ED 436 821 EA 030 147

AUTHOR Lee, Jaekyung
TITLE Assessing the Effectiveness of Instructional Resource Allocation and Use.
SPONS AGENCY National Center for Education Statistics (ED), Washington, DC.
PUB DATE 1999-04-00
NOTE 25p.; Paper presented at the Annual Meeting of the American Educational Research Association (Montreal, Quebec, Canada, April 19-23, 1999). Supported in part by the American Educational Finance Association.
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Achievement; Educational Assessment; Grade 8; Instructional Effectiveness; *Instructional Improvement; Junior High Schools; *Mathematics Instruction; *Resource Allocation; *School Effectiveness

ABSTRACT

This paper reports on a 1992 study of mathematics-assessment data by measuring key instructional resources and practices and by investigating the ways in which the resources and practices affect student learning in a multilayered, complex school system. The study examined research methods that assess the effectiveness of instructional resource allocation. The results encourage the possibility of applying objective measurement and multilevel analysis methods to survey and test data for assessing the effectiveness of instructional resource allocation and use. Findings show that the availability of both human and physical resources is positively associated with the level of desired instructional practices across states. Generally, the effect of human resources is greater than the effect of physical resources. Furthermore, the level of desired instructional practices is positively related to the level of academic achievement across states, although the relationship between instructional resources and practices varied from state to state. Setting desired levels of standards of instructional resources and practices may be tailored to individual states' unique status of resource allocation and use. States that are more effective in using physical resources than in using human resources should set standards for physical resources at higher levels than for human resources. (Contains 17 references.) (RJM)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 436 821

Assessing the Effectiveness of Instructional Resource Allocation and Use

Jaekyung Lee, Ph.D.

College of Education and Human Development

University of Maine

Paper presented at the Annual Meeting of the AERA (Montreal, Canada, April 20, 1999)

This research was supported in part by the American Educational Finance Association and National Center for Education Statistics under New Scholars Award Program. The views expressed herein are solely those of the author.

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

EA 030147

BEST COPY AVAILABLE

A large body of research, conducted over three decades following the Coleman report, has failed to find a systematic relationship between school resources and student achievement (Hanushek, 1997). The studies, so-called "education production function" studies relied on readily measurable indicators of school resources (i.e., per pupil expenditures, teacher salary, library resources) but failed to fully account for key aspects of schooling processes that affect student outcomes. On the other hand, another branch of research, so-called "effective schools" studies, found that desired instructional practices (i.e., clear goals and high expectations, opportunity to learn, monitoring student progress) enhance student achievement (Purkey and Smith, 1983; Lee, Bryk, and Smith, 1993). These case-studies sought to identify ellusive aspects of effective school context and process but failed to provide generalizable information on required resources as a sufficient base for policy making (Monk, 1992).

Need for filling such academic knowledge gap also comes from policy circles in which more state policymakers consider and adopt outcome-based school finance policies. This often involves efforts to set and enforce new standards for school resources and practices with an effective alignment with student outcomes. But the need is currently outrunning the knowledge base. It is challenging to collect valid and reliable data on instructional resources and practices as closely linked to student achievement. Researchers often utilize existing national databases that provide information on both schooling conditions and student achievement. For example, NAEP does not only assess students' academic achievement but also survey assessed students' teachers about instructional resources and practices in classrooms so that the teacher survey responses can be matched to the student test scores.

The most serious concern in research with NAEP data is one of errors of measurement and specification. In the case of teacher survey data, a question is raised about how to make sense of teachers' responses to multiple questions and to construct objective measures across teachers. Another question is how to choose appropriate unit of

analysis with the data collected through a multi-stage, complex sampling method and to examine multivariate relationships among several variables.

In light of these concerns, I conducted a more systematic analysis of the 1992 NAEP state mathematics assessment data by (1) objectively measuring key instructional resources and practices and (2) investigating the ways in which the resources and practices affect student learning in a multi-layered, complex school system. The study's objectives are to explore research methods for assessing the effectiveness of instructional resource allocation and use with the NAEP data and to draw policy implications for setting outcome-based standards of instructional resources and practices.

Research Design and Methods

In recognition of the potential provided by calculators and computers for increasing children's mathematical power, recommendations for improving math education often include more use of these tools in today's classrooms (NCTM, 1991). Instructional tools themselves, however, cannot develop a range of mathematical activities unless they are effectively used in classrooms. Improving teachers' knowledge and skills is essential in enhancing the quality of instructional services (Darling-Hammond, 1989; Shulman, 1987). Indeed, the current mathematics curriculum often fails to capitalize on the rich informal mathematics knowledge and understanding that children bring to instruction, and that school mathematics often seems divorced from such familiar activities (see Resnick, 1987; Romberg and Carpenter, 1986). To help anchor mathematics concepts for students, it is important to present mathematics in the "everyday" contexts and encourage students to work together in groups to solve problems. Thus, small-group work, using technologies, and problem solving in the context of projects can be considered positive signs of

implementation of many recent recommendations for the reform of school mathematics (see David, 1994; Weissglass, 1990; NCTM, 1991).

Building on the literature review, I develop an analytical framework to assess the effectiveness of instructional resource allocation and use. As shown in Figure 1, human and physical resources are allocated and used to deliver desired instruction, which in turn affects school performance. If schools manage to allocate and use more resources but fail to improve teaching and learning, the allocation and use of instructional resources is hardly effective. This raises two interrelated research questions. First, what kinds of instructional resources enhance quality instruction? Is school resource allocation effective? To explore those questions, I examined the relationship between instructional resources and practices (see arrows A and B in Figure 1). Secondly, what types of instructional practices boost student achievement? Is school resource use effective? To probe those questions, I further examined the relationship between instructional practices and school performance (see arrow C in Figure 1).

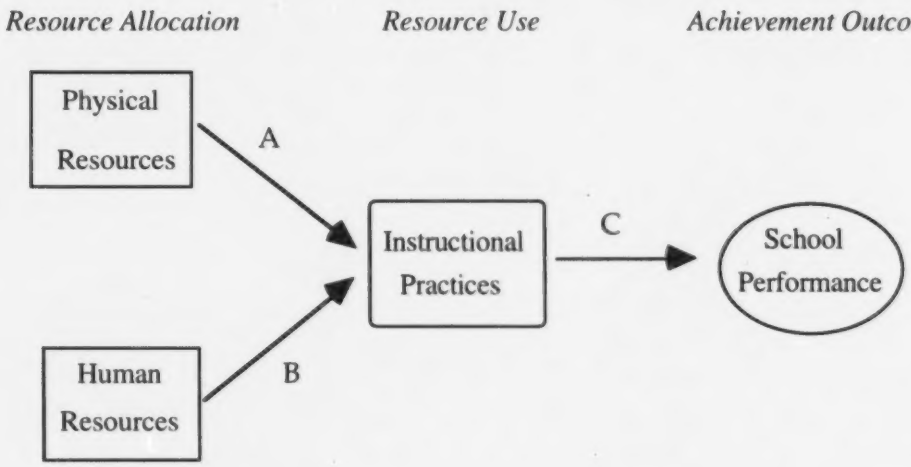


Figure 1. Analytical framework for assessing the effectiveness of instructional resource allocation and use

This study proceeds through two successive stages, that is, objective measurement and multilevel analysis for assessing the effectiveness of instructional resource allocation and use. Primary data sources are 1992 NAEP State Assessment in 8th grade math. While the math achievement of 8th grade students attending public schools in 41 states were assessed, information was also collected from the students' mathematics teachers about instructional materials and approaches currently used in their math classes. The first stage is to measure instructional resources and practices. The second stage is to link the measures to school performance. In the following sections, I explain the research methods employed at each of the two stages.

Objective Measurement Method

The first stage is to create objective measures of instructional resources and practices from the NAEP math teacher survey data. I chose to apply the item response theory (IRT) to measure the level of key instructional resources and practices. The basic idea of IRT theories and models is that from a set of observed responses to a set of items it is possible to derive measures or estimates of the underlying trait that have superior measurement and interpretive properties as compared to an unweighted sum of the item scores (Carroll, 1988). I chose to use the Rasch measurement model, among IRT models because the one-parameter Rasch model specifies only the position of an item on a difficulty scale and allows for more efficient analysis (see Wright and Stone, 1979; Wright and Masters, 1982).

The measurement of instructional resources and practices through an IRT method has theoretical grounds in this study. First of all, we need to make the scale of measurement linear. It is common practice in survey questionnaire analysis to compute differences between persons or groups in their raw scores (e.g., an unweighted sum or mean of the item scores) for their comparisons. Although such raw scores usually estimate

the order of person's location on a variable rather well, they never estimate the spacing satisfactorily. For example, the difference between score 10 and score 20 may not be the same as the difference between score 20 and score 30. This makes it difficult to compare schools or states on an interval scale in the level of resources and practices that their teachers reported. Further, it doesn't make sense to relate such nonlinear estimates of instructional resources and practices to the linear estimate of student achievement in NAEP assessment that is produced through an IRT scaling method.

Secondly, we need to make the scales of measurement for different types of inputs comparable given their cost differences. For example, hiring two new teachers does not cost the same as purchasing two new computers: their unit cost is different. The solution is to express both units into dollars: more expensive units will earn greater value. To measure different inputs on a common scale (like in dollar amount) from survey responses, we may regard item difficulty as reflecting the cost involved in each item (including not only financial but also human costs in acquiring or using those inputs for educational production). This allows us to express the measures of both human and physical resources on a common, difficulty-adjusted scale: being rated high on more difficult (i.e., probably most costly) items will get more credit.

Multilevel Analysis Method

The data collected under NAEP state assessment is hierarchical in nature because students are nested within schools, which in turn are nested within states. Multi-level analyses of the 1992 NAEP state assessment data involves examining the relations between instructional resources, practices, and outcomes through hierarchical linear models (Bryk and Raudenbush, 1992). The use of hierarchical linear modeling (HLM) on NAEP data will cope with the problem of sampling error resulting from the multi-stage sampling in NAEP. The measurement error resulting from the multiple imputation of NAEP scores will

be taken into account by averaging the parameter estimates obtained from the HLM analyses of five plausible values (Arnold, 1993).[1]

Multilevel analysis is also needed to capture interstate variations in the effectiveness of school resource allocation and use. The levels of school input and outcome as well as their relationships are presumed to vary substantially among the states. Figure 2 illustrates the hypothetical relationship between school input and outcome variables in two states, A and B.[2] State A does not only produce more outcome than state B at a given level of input but also has stronger relationship between school input and outcome. Thus, resource allocation and use in state A is regarded as more effective than state B.

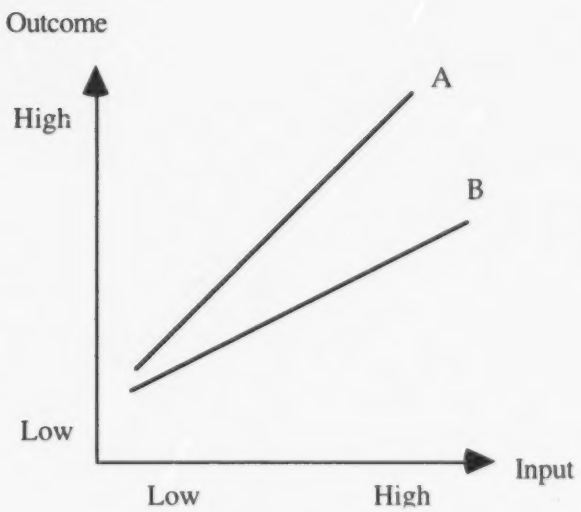


Figure 2. Hypothetical Relationship between School Input and Outcome Variables in States A and B

[1] NAEP used item response theory (IRT) to estimate proficiency scores in math for each individual student. Five plausible values for each sampled student result from five random draws from the conditional distribution of proficiency scores for each student.

[2] Instructional practices variable is treated as an outcome variable when it is predicted by instructional resources as input variables. But at the same time, instructional practices variable is treated as an input variable when it is related to school performance as an outcome variable.

Data Analyses and Results

Objective Measurement of Key Instructional Resources and Practices

Information on the availability of basic instructional materials and tools for students as well as teachers are obtained from the responses of 11,247 8th grade math teachers to four items in the 1992 NAEP teacher questionnaire to measure "physical resources" (see Table 1). Secondly, information on both pre-service and in-service teacher training in math content knowledge and pedagogical skills are obtained from the responses of 11,290 8th grade math teachers to twelve items in the 1992 NAEP teacher questionnaire to measure "human resources" (see Table 1). Finally, the responses of 10,982 8th grade math teachers to thirteen items on current classroom activities in the 1992 NAEP teacher questionnaire are used to measure "progressive instruction" (see Table 1).

BIGSTEPS, the Rasch measurement program, is used to construct objective measures from the responses of 8th grade math teachers to the 1992 NAEP teacher survey items. Both teacher measure and item difficulty are calibrated on the same logit scale. Because the difficulties of human resource items are likely to differ from those of physical resource items, the scale for human resource measures is equated with the scale for physical resource measures.

Table 1. Items Used to Measure Instructional Resources and Practices from the 1992 NAEP 8th Grade Mathematics Teacher Survey

Physical Resource (PR) Items
[1] How well does you school provide resources? (get all, most, some, none)
[2] Student access to school-owned 4-function calculators? (yes or no)
[3] Student access to school-owned scientific calculators? (yes or no)
[4] Are computers available for your math class? (yes or no)

Human Resource (HR) Items
[1] Training in estimation? (yes or no)
[2] Training in math problem-solving? (yes or no)
[3] Training in use of manipulatives? (yes or no)
[4] Training in use of calculators? (yes or no)
[5] Training in students' math thinking? (yes or no)
[6] Training in number systems and numeration? (yes or no)
[7] Training in measurement in math? (yes or no)
[8] Training in geometry? (yes or no)
[9] Training in probability or statistics? (yes or no)
[10] Training in abstract or linear algebra? (yes or no)
[11] Training in calculus? (yes or no)
[12] Training in methods of middle-school math? (yes or no)
Progressive Instruction (PI) Items
[1] How much emphasis on reasoning/analysis? (heavy, moderate, little/no)
[2] How much emphasis on communicating math ideas? (heavy, moderate, little/no)
[3] How often do students work in small groups? (daily, weekly, monthly, never)
[4] How often do students use measurement/geometry? (daily, weekly, monthly, never)
[5] How often do students use calculators? (daily, weekly, monthly, never)
[6] How often do students use computers? (daily, weekly, monthly, never)
[7] How often do students write reports/do projects? (daily, weekly, monthly, never)
[8] How often do students write about problem-solving? (daily, weekly, monthly, never)
[9] How often do students discuss math with others? (daily, weekly, monthly, never)
[10] How often do students work real-life problems? (daily, weekly, monthly, never)
[11] How often do students make up math problems? (daily, weekly, monthly, never)
[12] How often assess students with written responses? (weekly, monthly, yearly, never)
[13] How often assess students w/ projects/portfolios? (weekly, monthly, yearly, never)

Note. Response categories for each question are shown in parenthesis.

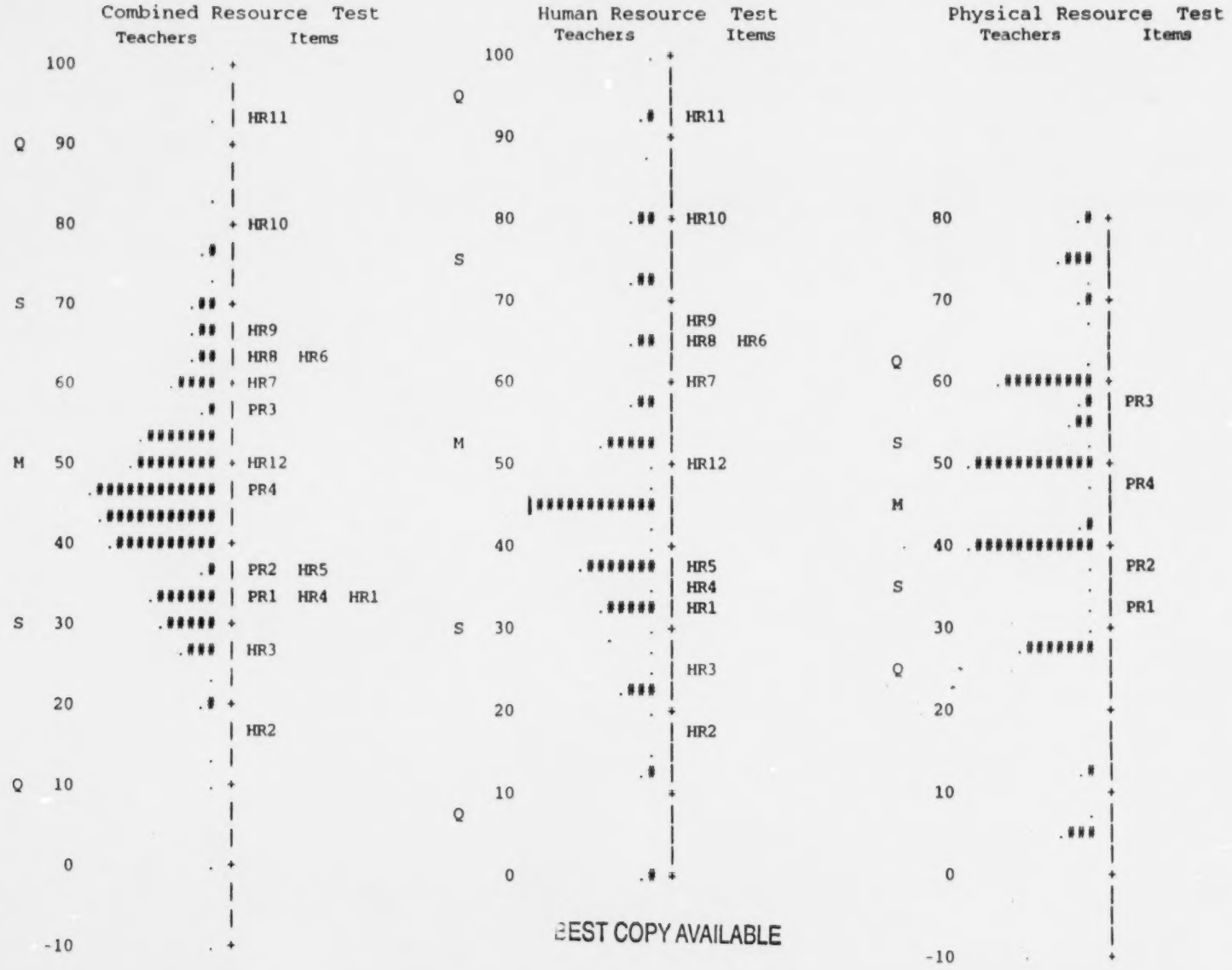
In Figure 3, the measures of those instructional resources are laid out vertically with the highest rating teachers and the most difficult items at the top. The item difficulty for instructional resource measures is scaled to have a mean of 50 with 10 units per logit. Two different sets of resource measures are obtained for each teacher by first jointly calibrating item difficulties with human and physical resource items together (see combined resource test in Figure 3) and then separately producing two different sets of teacher measures with item difficulties anchored on the combined calibrations (see human resource test and

physical resource test in Figure 3). As shown in Figure 3, human resource items are generally more difficult than physical resource items. This indicates that teachers experience greater difficulty in receiving professional training than in getting instructional materials.

Table 2 shows the results of calibrating progressive instruction items through the Rasch measurement method.[3] As with instructional resource measures, the item difficulty for instructional practice measure is scaled to have a mean of 50 with 10 units per logit. Items are hierarchically ordered in terms of their item difficulty to define a construct of progressive instruction. Goal-related items are less difficult than evaluation-related items, whereas the difficulty of practice-related items is dispersed according to the characteristics of the activity.

The difficulty of teachers' having students engage in a particular classroom activity seems to reflect the cost and complexity of implementing the activity: the more an activity requires expenses and efforts on the part of schools or teachers, the less likely teachers are to practice it. For example, having students write reports or do projects turned out to be the most difficult-to-practice. This can be explained by the fact that the activity incurs high opportunity cost by taking up most of the class time and thus reducing expected content coverage. Using computers is more difficult than using calculators because the former requires higher costs for purchase and greater complexity for operation than does the latter.

[3] Most instructional practice items included in this study involve four response categories asking teachers about the frequency of an instructional activity (1=never, 2=monthly, 3=weekly, 4=daily). A teacher who chooses the third category can be considered to have chosen "monthly" over "never" (first step taken) and also "weekly" over "monthly" (second step taken), but to have failed to choose "daily" over "weekly" (third step not taken).



BEST COPY AVAILABLE

Figure 3. Distributions of teacher measures and item difficulties. S and Q are placed one and two standard deviations, respectively, away from M, the mean of teacher measures. Item type: HR = human resources, PR= physical resources

Table 2. Summary Statistics of Rasch Measuremet: Progressive Instruction (PI) Items

Item No.	Goal (Emphasis)	Practice (Activity)	Evaluation (Assessment)	Measure (Error)	Misfit[a]	Point-Biserial[b]
7		Write reports/ Do projects		71.73 (.22)	.75	.39
6		Use Computers		63.40 (.16)	1.49	.13
11		Make up math Problems		61.53 (.15)	.96	.41
4		Use measurement		59.19 (.14)	.72	.38
13			Projects/ Portfolios	57.50 (.14)	.94	.44
8		Write about problem-solving		56.12 (.13)	.85	.50
12			Written responses	50.07 (.12)	1.11	.43
3		Work in small groups		44.63 (.12)	.87	.45
5		Use calculators		42.79 (.12)	1.60	.25
10		Work real-life math problems		38.85 (.12)	.74	.44
2	Communicating math ideas			36.25 (.16)	.93	.41
9		Discuss math with others		34.21 (.13)	.99	.42
1	Reasoning/ Analysis			33.73 (.17)	.96	.37

Note. Items are arranged and shown in difficulty order.

[a] Values substantially greater or less than 1 indicate that items poorly define the construct.

[b] The coefficient indicates a correlation between the teachers' responses to an item and their total scores (i.e., progressive instruction measure).

Multilevel Analysis of Instructional Resource Allocation and Use

The purpose of drawing teacher samples in the NAEP data was not to estimate the attributes of the teacher population, but to correlate student performance with the characteristics of their teachers (Johnson et al., 1994. The NAEP 1992 Technical Report, p. 86). Thus, teacher measures as defined and constructed in the previous section are matched to their students for examining their relationship with student outcomes. Since this study focuses on schools as primary unit of analysis, I produced the school average measures of instructional resources, practices, and math achievement. It is presumed that the relationships between the three school variables vary among states. The HLM/2L program is used to partition the total variance in outcome variable into its between-school and between-state components. First, using a sample of schools from each state (3,544 schools in 40 states), a school-level linear regression model is estimated for each state to identify the association of input variable(s) with an outcome variable. Simultaneously, a state-level regression model is estimated across 40 states to examine interstate variations in the mean level of outcome (intercept) and the input-outcome relationship (slope).

Table 3 summarizes the results of the HLM analysis on the relationship between instructional resources (input) and practices (outcome). The effect of human resources on progressive instruction (HR effect) is .135, whereas the effect of physical resources on progressive instruction (PR effect) is .089. All these mean effects include adjustment for the other variable in the model, and all are statistically significant at probability levels less than .001. Further, the difference in effect size between these two types of resources (i.e., .135 - .089 = .064) is also statistically significant (reject H_0: .135 = .089 with chi-square statistic of 5.107, df=1, $P < .05$). In other words, human resources are generally more cost-effective than physical resources in producing progressive instruction in math. This indicates that the current school delivery of progressive instruction is labor-intensive.

Table 3. Results of HLM Analyses: Effects of Human Resources (HR) and Physical Resources (PR) on Progressive Instruction (PI)

Estimated Effects				
	Beta Coefficients	Standard Error	t-Statistic	p-Value
Intercept (Mean PI)	44.655	.269	166.17	.000
Human Resources (HR Effect)	.135	.014	9.583	.000
Physical Resources (PR Effect)	.089	.010	8.574	.000

The Chi-Square Table				
Parameter	Estimated Variance	Degrees of Freedom	Chi-Square	p-Value
Mean PI	2.184	39	277.77	.000
HR Effect	.005	39	144.23	.000
PR Effect	.002	39	66.94	.004

Correlations among Random Effects		
	Mean PI	HR Effect
HR Effect	-.153	
PR Effect	.555	-.546

Reliability of Random Effects	
Mean PI	= .736
HR Effect	= .634
PR Effect	= .386

The correlations among the random effects indicate the general structure of instructional resource allocation. A high level of progressive instruction is associated with a smaller HR effect (r = -.153) and a greater PR effect (r = .555). This indicates that states producing more progressive instruction (i.e., more frequent student-centered, higher-order learning activities using modern technologies) tend to use physical resources more effectively than human resources (i.e., PR-intensive or HR-saving). There is also a substantial negative correlation between HR effect and PR effect (r = -.546). This indicates

that states using physical resources more effectively tend to use human resources less effectively.

Table 4 summarizes the results of HLM analyses on the relationship between progressive instruction (input) and school performance (outcome).[4] School average measure of instructional practices is significantly positively related to school average math achievement score. This indicates that an effective use of instructional resources involves more frequent student-centered, higher-order learning activities with use of modern technologies, and thus leads to an improvement of school performance.

Table 4. Results of HLM Analyses: Effect of Progressive Instruction (PI) on School Performance (SP)

Estimated Effects				
	Beta Coefficients	Standard Error	t-Statistic	p-Value
Intercept (Mean SP)	266.671	1.621	165.554	.000
Progressive Instruction (PI Effect)	.260	.064	4.037	.000

The Chi-Square Table				
Parameter	Estimated Variance	Degrees of Freedom	Chi-Square	p-Value
Mean SP	99.682	39	1129.78	.000
PI Effect	.081	39	89.10	.000

Correlation between Random Effects	
	Mean SP
PI Effect	.268

Reliability of Random Effects	
Mean SP	= .949
PI Effect	= .485

[4] The parameter estimates from the HLM analyses are based on the average parameter estimates from separate HLM analyses of the five plausible values.

The effect of progressive instruction on school performance turned out to vary significantly among the states. In other words, some states are better able to link instructional practices to school effectiveness. However, higher performing states do not show a stronger relationship between progressive instruction and school performance (i.e., more effective resource use); correlation between the random effects is positive but very low (r=.268).

Discussion

Clearly, there are limits to the inferences that can be drawn from teacher responses to a survey questionnaire. One may doubt the idea of measuring different types of resources on the same scale with an adjustment for their potential cost differences. Further, the cross-sectional nature of NAEP data limits causal inferences that can be made about the relationships among school resources, practices, and outcomes. The analysis of differences between high and low performing schools in their instructional resources and practices helps us identify the correlates of schol performance, but does not allow us to determine any causal direction of the relationship. In other words, the central question remains: do more resources and better practices lead schools to higher performance or do simply higher performing schools draw better teachers and get more resources? Thus, the findings of this research should be interpreted with caution.

Nevertheless, this exploratory study sheds light on the possibility of applying objective measurement and multilevel analysis methods to survey and test data for assessing the effectiveness of instructional resource allocation and use. Several patterns of instructional resource allocation and use emerge from the analyses of the 1992 NAEP state 8th grade math teacher survey and student assessment datasets. First, the availability of both human and physical resources is positively associated with the level of desired

instructional practices across states. Generally, the effect of human resources is greater than the effect of physical resources. States that produce more progressive instruction (i.e., student-centered, higher-order learning practices with use of technology) tend to use physical resources more effectively than they use human resources (i.e., more capital-intensive or labor-saving). Second, the level of desired instructional practices is positively related to the level of academic achievement across states. While states vary substantially in the relationship between progressive instruction and school performance, states that perform at a higher level are not necessarily more effective in instructional resource use.

Implications for Setting Standards of Instructional Resources and Practices

In the midst of keen policy interest in standards-based education reform, the findings of this study has implications for setting outcome-based standards of instructional practices and resources. When assessments are used for certification of teachers or for determining the level of instructional expenditures, the need for explicit standards is inevitable. In order to align standards of classroom resources and practices with student performance standards, the measures of those resources and practices should be valid and reliable enough for meaningful interpretation.

There may be two different approaches to standards-setting as we transform verbal descriptions of standards into cut scores (i.e., the numeric values that operationalize "how good is good enough"). One approach is setting standards of inputs independently from outcomes. If we collected data on instructional resources and practices through either teacher survey or assessment, we could use the Angoff method to set benchmarks on the scales of those resources and practices. In the Angoff procedure, judges are asked to imagine a group of teachers at the threshold of a given input standard and estimate the

probability of giving a keyed response to each item.[5] The average probability estimate over judges is defined as the item minimum pass level (MPL), and the sum of the item MPLs becomes the passing score (Kane, 1994). This approach to standard-setting for educational inputs has a potential problem in that such stand-alone input standard has no link to an outcome standard so that it may be too high or too low to meet a desired outcome level.

In light of these problems, an alternative approach that I would suggest is to directly link input standards to a pre-existing outcome standard based on their empirical relationship. This will allow us to pinpoint the location of an input variable as corresponding to a desired outcome standard, convert the cut score into individual item response probabilities, and interpret them in narrative, integrative ways. Thus, this approach follows the reverse procedure (i.e., proceeding from cut score to verbal description) of what is taken for setting student performance standards. For an illustration, the following equation is derived from the estimated relationship between progressive instruction (X) and math achievement (Y) across states: 266.7 is the grand mean of Y; .26 is the estimated effect of X on Y; 44.5 is the grand mean of X.

$$Y = 266.7 + .26 * (X - 44.5)$$

Suppose that we want to identify the level of progressive instruction that corresponds to the "Basic" achievement level in eighth grade mathematics as defined by the National Assessment Governing Board. The Basic level was set at a score of 262 on 0 to 500 scale, and eighth-grade students performing at this level should exhibit evidence of conceptual and procedural understanding in the five NAEP content strands. Then, the measure of progressive instruction as corresponding to the Basic achievement level (cut score of 262) is 26.4. This level of progressive instruction can be interpreted in

[5] In practice, judges would estimate what percentage of the group would answer the item correctly in the case of assessment-type data or estimate what percentage of the group would give positive rating in the case of survey-type data.

probabilistic terms based on the gaps between the practice measure and the difficulties of items below (gap shown in one tenth of logit unit).

- The probability of having students write reports or do projects daily or weekly is only 1 percent (gap = -45.3).
- The probability of assessing students with projects/portfolios weekly or monthly is about 5 percent (gap = -31.1).
- The probability of having students work real-life math problems daily or weekly is about 25 percent (gap = -12.45).
- The probability of giving a heavy emphasis on reasoning/analysis is about 30 percent (gap = -7.3)

The above desriptions of selected item responses indicate that instructional practices corresponding to the "Basic" achievement level are hardly progressive: the overall percentage of students at the Basic level who have opportunities to get involved in regular progressive learning activities with a strong emphasis on higher-order thinking is even less than 50 percent.

Likewise, we can identify the level of progressive instruction that matches the "Proficient level" of 8th grade mathematics achievement. The Proficient cut score is 299, and students performing at this level should apply math concepts and procedures consistently to complex problems in the five NAEP content strands. The measure of progressive instruction that matches the achievement score of 299 is 168.7, which indicates that teachers regularly practice all of the desired classroom activities with 100 percent certainty. Such extraordinary level of progressive instruction, far beyond the distribution of sample schools, may be required for schools to perform at the Proficient level on average. Nevertheless, it is difficult to extrapolate the regression line to identify the value of X associated with mean Y, because the predictive ability of the regression line falls markedly

as X departs progressively from the mean of X. Thus, it is more reasonable to set performance levels near systemwide mean achievement scores so that a fixed predictor value corresponding to the conditional mean outcome value can be identified with greater accuracy.

Once we have set standards for instructional practices, the same procedures can be applied to setting standards for instructional resources. Here another question is raised as to setting standards for multiple inputs that are simultaneously linked to one common outcome. Suppose we run a multiple regression of Y on several Xs and identify the unique (partial) effect of each X on Y. If the input variables were measured on a common scale with an adjustment for their probable cost differences, unstandardized regression (slope) coefficients could be used as the indicators of cost-effectiveness.[6] Then, the coefficient becomes a weight for each X in determining the level of each X required for producing a certain level of Y: the more cost-effective X is, the more it should be used. To illustrate this idea, the following equation is derived from the estimated relationships of human resources (X1) and physical resources (X2) with progressive instruction (Y) across states: 44.6 is the grand mean of Y; .14 and .09 each are the estimated effects of X1 and X2 on Y; 47.0 and 46.3 each are the grand means of X1 and X2.

$$Y = 44.66 + .14 * (X1 - 47.0) + .09 * (X2 - 46.3)$$

Since the effect of X1 on Y1 is about 1.5 times greater than the effect of X2 on Y, the standard for X1 should be also 1.5 times higher than standard for X2. When we plug (1.5 * X2) into X1 for substitution, the above equation is simplified as follows:

$$Y = 33.91 + .3 * (X2)$$

---

[6] This procedure is different from the conventional method that uses standardized multiple regression coefficients as the basis of determining the effect sizes of input variables that have different scales.

If a desired level of Y is 44.53 (the grand mean of progressive instruction), then we get an X2 of 35.4 and an X1 of 53.1. Consequently, the standard of human resources should be set at the level 1.5 times higher than the standard of physical resources.

Despite the aggregate pattern of resource allocation and use across states, it needs to be noted that the relationship between instructional resources and practices was found to vary from state to state. This means that setting desired levels of standards of instructional resources and practices may be tailored to individual states' unique status of resource allocation and use. For instance, states in which schools are found to be more effective in using physical resources than in using human resources should set standards for physical resources at higher levels than for human resources so that both kinds of resources are allocated and used more cost-effectively to meet desired levels of instructional practices and outcomes.

References

Arnold, C. A. (1993) Using HLM with NAEP. Unpublished paper presented at the Advanced Studies Seminar on the Use of NAEP Data for Research and Policy Discussion, Washington, D.C.

Bryk, A. S., and Raudenbush, S. W. (1992) Hierarchical Linear Models. Newbury Park: Sage Publication.

Carroll, J. B. (1988). Future developments in educational measurement. In J. P. Keeves (Ed.) Educational research, methodology, and measurement: An international handbook. New York: Pergamon Press.

Darling-Hammond, L. (1989). Teacher professionalism: Why and how?," in A. Lieberman (ed.) Schools as Collaborative Cultures: Creating the Future Now. London: Falmer Press.

David, J. L. (1994) Realizing the promise of technology: A policy perspective. In Technology and Education Reform: The Reality Behind the Promise, ed. B. Means. San Francisco: Jossey-Bass.

Hanushek, E. A. (1997). Assessing the effects of school resources on student performance: An update. Educational Evaluation and Policy Analysis, 19, 141-164.

Kane, M. (1994). Developing standards for the National Assessment of Educational Progress. Proceedings of joint conference on standard setting for large-scale assessments. Washington, DC: U.S. GPO.

Lee, V., Bryk, A., & Smith, J. (1993). The organization of effective secondary schools. Review of Research in Education 19, 171-267.

Monk, D. H. (1992). Education productivity research: An update and assessment of its role in education finance reform. Educational Evaluation and Policy Analysis. 14(4), 307-32.

National Council of Teachers of Mathematics. (1990) Professional Standards for Teaching Mathematics. Reston, VA: NCTM.

Purkey, S. D. and M.S. Smith. (1983). Effective schools: A review. Elementary School Journal 83, 335-351.

Resnick, L. B. (1987) Education and Learning to Think. Washington, D.C.: National Academy Press.

Romberg, T. A. and Carpenter, T. P. (1986) Research on teaching and learning mathematics: Two disciplines of scientific inquiry. In Handbook of Research on Teaching, ed. M.C. Wittrock, New York, NY: Macmillan.

Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. Harvard Educational Review 57, 1-22.

Weissglass, J. (1990). Cooperative learning using a small-group laboratory approach. In Cooperative Learning in Mathematics. ed. N. Davidson, Menlo Park, CA: Addison-Wesley Publishing Company.

Wright, B. D. & Masters, G. N. (1982) Rating Scale Analysis. Chicago: MESA Press.

Wright, B. D. and Stone, M. H. (1979) Best Test Design: Rasch Measurement. University of Chicago: MESA Press.